



Published in final edited form as:

J Res Pract. 2018 ; 14(1): .

Conducting a Large Public Health Data Collection Project in Uganda: Methods, Tools, and Lessons Learned

Bert Stover,

School of Public Health, University of Washington, Seattle, UNITED STATES

Flavia Lubega,

College of Health Science, Makerere University, Kampala, UGANDA

Aidah Namubiru,

College of Health Science, Makerere University, Kampala, UGANDA

Evelyn Bakengesa,

College of Health Science, Makerere University, Kampala, UGANDA

Samuel Abimerech Luboga,

College of Health Science, Makerere University, Kampala, UGANDA

Frederick Makumbi,

College of Health Science, Makerere University, Kampala, UGANDA

Noah Kiwanuka,

College of Health Science, Makerere University, Kampala, UGANDA

Assay Ndizihiwe,

Centers for Disease Control and Prevention, Entebbe, UGANDA

Eddie Mukooyo,

The Resource Centre, Ministry of Health, UGANDA

Erin Hurley,

Centers for Disease Control and Prevention, UNITED STATES

Travis Lim,

Centers for Disease Control and Prevention, UNITED STATES

Nagesh N. Borse,

Centers for Disease Control and Prevention, UNITED STATES

James Bernhardt,

School of Public Health, University of Washington, Seattle, UNITED STATES

Angela Wood,

School of Public Health, University of Washington, Seattle, UNITED STATES

Lianne Sheppard,

School of Public Health, University of Washington, Seattle, UNITED STATES

Scott Barnhart, and

School of Public Health, University of Washington, Seattle, UNITED STATES

Amy Hagopian

School of Public Health, University of Washington, Seattle, UNITED STATES

Abstract

We report on the implementation experience of carrying out data collection and other activities for a public health evaluation study on whether U.S. President's Emergency Plan for AIDS Relief (PEPFAR) investment improved utilization of health services and health system strengthening in Uganda. The retrospective study period focused on the PEPFAR scale-up, from mid-2005 through mid-2011, a period of expansion of PEPFAR programming and health services. We visited 315 health care facilities in Uganda in 2011 and 2012 to collect routine health management information system data forms, as well as to conduct interviews with health system leaders. An earlier phase of this research project collected data from all 112 health district headquarters, reported elsewhere. This article describes the lessons learned from collecting data from health care facilities, project management, useful technologies, and mistakes. We used several new technologies to facilitate data collection, including portable document scanners, smartphones, and web-based data collection, along with older but reliable technologies such as car batteries for power, folding tables to create space, and letters of introduction from appropriate authorities to create entrée. Research in limited-resource settings requires an approach that values the skills and talents of local people, institutions and government agencies, and a tolerance for the unexpected. The development of personal relationships was key to the success of the project. We observed that capacity building activities were repaid many fold, especially in data management and technology.

Index Terms

health services research; public health evaluation; research context; data collection method; limited-resource setting; research implementation; project management; research partnership; health system strengthening

1. Introduction

This article reports on methods and lessons learned from implementing a large nationwide data collection effort in Uganda. Similar to other large projects in limited-resource settings that require mobilizing teams of individuals (such as immunization campaigns), our research effort was challenging and fraught with unexpected problems.

The US Centers for Disease Control and Prevention (CDC) entered into a cooperative agreement with the University of Washington (UW) in 2010 to study the effects of the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) on recipient country health systems, in search of both positive and negative effects, using Uganda as the case example. Effects were measured with health service utilization and health system strengthening indicators. The UW in turn, contracted with Makerere University, Uganda and the Uganda Ministry of Health to implement data collection for the study in coordination with all partners. This retrospective study focused on the PEPFAR scale-up, mid-2005 through mid-2011 to provide detailed information for consideration in policy discussions and future PEPFAR efforts. Results of the study are reported elsewhere (Luboga et al., 2016).

Several factors led to the selection of Uganda as the study setting, including the availability of the government's Health Management Information System (HMIS), the interest of the Ministry in conducting the study, the intensity and relative exclusivity of PEPFAR funding, and the variable distribution of PEPFAR funding across the country (allowing comparison of funded with non-funded sites) (Gladwin, Dixon, & Wilson, 2003). Uganda's HIV prevalence rate remained above 6 per cent among adults, providing a continuation of the HIV burden throughout the observation period, 2005-2011 (Uganda Bureau of Statistics & ICF International, 2012).

To simultaneously deploy 6 teams of 3 people across the country to collect data from 315 facilities, we developed creative approaches to deal with the multiple technical and operational challenges we faced. We found the most important thing was to sustain collaborative partnerships, despite multiple opaque institutional bureaucracies and multiple personnel changes. This required interpersonal and management skills, humor, creativity, organization, experience, and cultural sensitivity. To build these valuable partnerships, we spent time in the field together piloting data collection, hosted appropriate social gatherings, exchanged presentations with partner institutions, communicated frequently by phone across multiple continents, and identified new projects to pursue when this one was scheduled to be over.

We also worked to build research capacity to advance Uganda's health system, so we were intentional about including early-career young professionals at all steps in the project. Our teams expanded their skills in multiple areas they can use in other projects, especially in data management and technology.

2. Preparatory Steps

2.1. Planning, Contracting, and Budgeting

A research protocol was developed by the Health Systems and Human Resources Team within the Health Economics, Systems and Integration Branch in the Division of Global HIV/AIDS at the US Centers for Disease Control and Prevention (CDC). The CDC entered into a cooperative agreement with the University of Washington (UW) in September 2010, allocating USD 1,170,588 over 3 years. The UW, in turn, contracted with its long-term partner, Makerere University, Uganda, and structured its contract based on deliverables, such as the number of facilities from which data were to be collected and the number of manuscripts to be drafted. A deliverables-based contract (rather than paying on budget line-items for individual cost elements such as staffing, supplies, or travel) conveyed our trust in the Ugandan partner to deliver study products with little direct supervision by the UW.

Early in the project, the collaborators developed a list of likely manuscripts to emerge from the project, with a lead author identified for each manuscript. Our experience was that researchers with first-author responsibility would ensure the required information was collected in relation to their manuscripts throughout the planning, data collection, and analysis processes. Detailed notes of our twice-monthly conference calls and trip reports were kept to track team decisions and progress. We established a password-protected project website to store materials such as meeting minutes, data variables, data dictionaries, analysis

plan, links to other source material, training materials, questionnaires, abstracts, bibliographies, and manuscripts.

The costs for data collection logistics were approximately USD 1,000 per site. Makerere budgeted for data collection by calculating days of data collection required per site, number of personnel, per diems for meals and overnight accommodations, mileage estimates, contracts with companies to provide vehicles and drivers, office supplies, and accounting services, among other costs. Makerere provided the local knowledge of banking procedures and exchange rate patterns, while maintaining accounting systems for paying field teams. To maintain cash flow, Makerere budgeted time for clearance of payments through multiple organizations, international transfer of funds, and transfer of funds to the field.

However, the negative effects of the global financial crisis, fluctuating exchange rates, and the brief U.S. government shutdown (during October 2013) were significant. The shutdown resulted in the cancelation of supplemental budgets to fund parts of this project, and disrupted the lives of field staff who had committed to the work. These factors created a hyper-vigilant approach to expenses, disruption to the research teams work schedules, and delays in data collection efforts.

2.2. Addressing Legal Issues: Data User Agreement and Human Subject Protection

We proposed and drafted a formal data user agreement, believing it could prevent confusion later about data ownership and permissions to publish manuscripts based on data collected. The agreement includes the names and positions of principals from each organization, duration of the agreement, requirements for approving analysis and publication, data ownership, data protection, dispute resolution, and conditions for termination of the agreement. This was a unique arrangement between the Ministry, partners, and donors. While we drafted this agreement at the beginning of the study, the final agreement was not signed by all parties until the end of 2012 (2 years into the study). The Ministry of Justice signed the data user agreement on behalf of the Government of Uganda.

Multiple institutional human subjects review boards approved the study before we began data collection, including those at the UW, Makerere University, Uganda National Council of Science and Technology, and CDC. The Ugandan Ministry of Health provided legal access to the government's Health Management Information System (HMIS) forms, as well as a letter of introduction from a recognized official, to facilitate access to health services data. The health services data collected do not contain personally identifiable information.

2.3. Determining Data Sources and Preparing Questionnaires

This study had two rounds of data collection, the first from the compiled district-wide summary reports at the 112 district offices, and the second from the 315 health care facilities directly. This article describes only the data collection efforts from the health care facilities, because district office data collection was done differently.

We used both Ugandan government and U.S. plan-specific data systems (i.e., specific to the President's Emergency Plan for AIDS Relief [PEPFAR]), along with additional sources to provide control data for statistical modeling. Uganda's HMIS was launched by the Ministry

of Health in 2005, coinciding with the start of our observation period, 2005–2011. The HMIS consisted of several routine forms that health facility personnel were expected to complete both monthly and annually, for all public and most private facilities. We conducted pilot field visits to determine whether the data would be sufficiently accurate and available for our purposes. Older data tended to be less well kept, filing space was uniformly inadequate and some forms had been damaged by the elements. HMIS data were not available electronically at district or national levels.

Beginning in 2004, PEPFAR (through USAID) funded a private U.S. contractor to establish the Monitoring and Evaluation of Emergency Plan Progress data system in Uganda. Data variables collected by this contractor included counts of HIV services, including the year and health facility where services were provided. We visited the contractor several times to ensure we understood the meaning of each variable, and collected the data set electronically. After considering alternatives, we used the counts of PEPFAR-provided anti-retroviral patient services to serve as the primary measure of PEPFAR investment.

To control for determinants of health and other variables that could confound the effects of PEPFAR investment, our models included prevalence of latrine sanitation, and the proportion of primary school enrolment from the Ugandan Bureau of Statistics. Obtaining population data at the district level for each year of the study was difficult, partly due to the splitting of districts during the observation period, 2005–2011. Determining dates of district splitting was complicated by conflicting source information and uncertainty of official start date versus contributions of HMIS data. We used data from the Ugandan AIDS Indicator Survey for 2006 and 2011 to provide measures of HIV prevalence, and the Uganda Demographic and Health Survey for 2004/2005 and 2011 to provide measures of underweight children, under-5 mortality rates at the region level (Ministry of Health Uganda, ICF International, CDC Uganda, USAID Uganda, & WHO Uganda, 2012; Uganda Bureau of Statistics & ICF International, 2012).

To ensure we had the necessary variables to answer the research questions in our protocol, we created a table containing each specific study aim, the HMIS variables we thought would best supply information to address the aim, and any questions we would ask the health system administrators to supplement the HMIS variables. We asked facility administrators to reflect on changes in health system indicators, including volume and availability of health services, over the observation period, 2005–2011. We also sought administrators' opinions about the effect of PEPFAR in each element of our conceptual framework of the health system. We collected both quantitative and qualitative data.

Some ways of asking questions did not produce the data we expected, in part because of cultural perspectives and language differences. Therefore, our team changed the original data collection instruments to obtain responses more useful for analysis. We sought to ensure that the questions required a response, were not conditional, asked for a single response, did not duplicate questions asked elsewhere, were forced choices when appropriate, and elicited responses that were analyzable. We included open-ended questions to elicit opinions and observations and coded the text responses using qualitative software *Atlas.ti* (Lohman et al., 2017).

2.4. Pilot Testing

We spent about 6 months conducting multiple facility office visits to pilot our methods and instruments. Our experiences of piloting data collection taught us many things. Some information listed in the HMIS forms was not available for the early years of PEPFAR scale-up. To evaluate data accuracy, we collected data from key variables in the HMIS and reviewed the results with similar data from health facility log books and Ugandan health services researchers.

From our Ugandan partners and field experience we learned that, for our teams to be welcomed, visits needed careful advance communication. Established cultural norms dictated that we present ourselves and our task to the facility administrator and other key personnel upon entry, and that training our teams to communicate well during these interactions was important. Because we were asking HMIS clerks to contribute their time to our study, we decided to provide a modest compensation to the facility. HMIS clerks often spent 4 or more hours with the team to welcome and orient them to the data storage facilities. Facility administrators were uniformly gracious in offering tea while we presented study details, described our collaborative partnerships, and provided evidence of permission to access medical data. Piloting helped us learn and practice these culturally expected behaviors and aided our team in planning time and resources for supporting the data collection teams in the field.

2.5. Recruiting Data Collectors

The proliferation of PEPFAR-related research in Africa has resulted in a skilled pool of recent graduates and junior faculty available for these intermittent assignments. Makerere University principals recruited data collection staff by circulating e-mail announcements through their networks. Applicants responded with cover letters and résumés. We selected applicants who demonstrated the skills we needed, basic knowledge of the health system, interest in the study, availability, and motivation. We screened out those with too many competing demands on their time and who had strong opinions about particular ways of doing certain aspects of the study. We believed that those with strong opinions about how they would conduct the research differently would find it difficult to apply the study guidelines and methods consistently.

2.6. Team Organization and Roles

We assembled teams with an eye on balancing social and data collection skills, geographical knowledge and preference, and gender distribution. During the training week, we had an opportunity to get to know our data collectors through exercises, question and answer sessions, simulations, and interactions. Team assignments were made at the end of the training.

Each team comprised two data collectors, a data entry person, and a driver (the latter provided by the vehicle vendor), and was assigned one of six geographic regions. Teams were asked to manage their own travel schedules and routes, including rest periods. We trained more team members than were expected to be in the field so we had substitutes

available for emergencies. We rotated people to different teams or roles during the several-month data collection period to provide rest, and to balance work type and environment.

Team roles were flexible, but typically one team member interviewed the facility manager, completed the questionnaire, and approached pharmacy staff for drug stock-out information. The second team member worked with the HMIS staff person to collect forms, and scan and upload these. The third team member remained at the Makerere University data center to receive, track, evaluate the quality of, and enter data as they were posted to the web server from the field. We thought it was important for each field team to have a specific data entry team member back at Makerere to ensure consistent transfer of data and individual accountability. Drivers were responsible for vehicle maintenance, gas, determining routes, security, and being consistently available for transportation as needed.

Senior Makerere faculty members oversaw data collection, and a team of skilled administrative assistants saw to organizational and fiscal details. A Makerere-based technician was assigned to help resolve electronic hardware, software, and connectivity problems.

While payment was initially offered to team members based on the number of days they spent in the field, we were obliged to change to paying on the basis of facilities visited (and data entered), with some allowances for far-flung locations. This performance-based payment method was a response to a very tight budget and our needs for efficiency. Prior to implementing the performance-based method we were concerned there might be a loss of attention to data accuracy and completeness due to data collectors' focus on completing facility visits and data entry. On comparison of accuracy and completeness of data before and after this change, we concluded there was no obvious pattern that could be linked to the new payment method.

2.7. Data Collector Training and Research Capacity Building

We provided 5 days of training at a Kampala hotel for 26 data collectors and data entry staff, including extra trained staff for substitution as needed. We took an individual photo of each trainee to create a photo gallery to help people get to know each other, and to provide each team member with a photo ID lanyard.

We developed a comprehensive field manual. The manual provided a structure and focus for training, and was updated as issues were worked through and new solutions adopted by the team. An outline of the manual is presented in Table 1. The manual and data collection instruments were made available online (at <http://hdl.handle.net/1773/41868>). Trainees reported they were more likely to read the sections of the manual after the content was discussed in training. Other large-scale survey projects have found a field manual was key to the consistency of collected data and field workers depended on the manual for following study protocol (Mitti, 2014, Hagopian, et al., 2013).

Consistent with the U.S. Global Health Initiative principles, including capacity-building, we sought to build research skills among junior faculty and new graduates who served as data collectors, data entry staff, and analysts (U.S. State Department, 2012). For example, during

our welcome meeting with new staff, we introduced the scientific goals and methods of the project to provide a theory-based background and develop a familiarity with the idea of the research, and its implications. During the training week, we invited our data collectors to help improve the questionnaire by testing it on each other and then going to a testing site in the field for live piloting. We offered different versions of questions and asked our team to assess whether questions were clear and sensible, and if the information was regularly available in the field during the pilot. We found the experience of collaboratively creating, implementing, and revising the data management system helped to develop skills useful for research practice. Indeed, other researchers conducting surveys in Sub-Saharan Africa have found that, when done well, survey data collection can lead to capacity building (Carletto, 2014).

In the training week, data collectors practiced entering and verifying data, and uploading it to the data center using the specific naming conventions required for file identification and management. During pilot testing in the field, each team went to a health facility and completed all steps—including meeting and negotiating with facility staff, using car batteries for powering equipment, enlisting the cooperation of the medical records personnel, scheduling and interviewing the facility administrator, and obtaining internet access and uploading data to our data center.

We employed hands-on exercises during training to demonstrate data collection and entry on both laptop data entry screens and smartphones. Teams practiced the scanning of HMIS forms and paper questionnaires, with the goal to be sure everyone understood how to upload data to the correct folder and filename. Each team member, including project organizers, signed a data quality and integrity contract to ensure everyone was committed to collecting high quality data. On the fourth day of the training week, we deployed each team with its car and driver to a nearby facility to practice live data collection and uploading of all data. We debriefed this experience and made revisions to procedures on the last day of the training week.

2.8. Providing Field Supplies and Equipment

Each field team was equipped with a laptop computer, a small portable internet modem, and a smartphone. A power inverter was provided for backup to operate electronics from the 12 Volt battery in the cars. We implemented password protection on computers, supplied a backup operating system on a hard drive to recover from possible computer crashes, and installed high quality antivirus software.

We provided large tubs for each car that included the field manual, checklists and clipboards, office supplies, notebooks for fieldnotes, questionnaires, copies of the HMIS forms, letters of introduction, maps, lists of facilities with contact information, and DVDs for backing up data from each facility. Materials for financial management included a locking cashbox, a ledger, and a receipt book. Each member had a laminated photo identification card on a lanyard, and business cards to distribute to facility personnel. We also provided each team a portable folding table and chair, because the space for data collection was not always available and we wanted to limit any disruption of normal work at the health facility. We also collected emergency contact information for each field team member.

2.9. Specifying the Use of Smartphones

Field teams collected a few limited data points on smartphones using Open Data Kit (ODK) Android-based software. ODK software was developed and is supported by the UW, and is free for research use (OpenDataKit, n.d.). Phones allowed us real-time automatic capture of global positioning system readings at healthcare facilities, a photo of the facility sign with team members, and time of arrival. We asked teams to complete questions concerning team management information on the phones, including road conditions, availability of facility personnel, and if follow-up visits were required. These data were collected on “Facility Arrival” and “Facility Exit” ODK forms, and variables are detailed in the field manual available online (at <http://hdl.handle.net/1773/41868>). If phone service coverage permitted internet connection, data were to be uploaded immediately.

We considered using smartphones to record data from the administrator interview. While we developed and tested several ODK format forms, and acknowledged their advantages over paper forms, we accepted the teams’ preference for the paper forms because of the hassles involved in extended text entry on small format phones with miniature keyboards.

3. Data Collection in the Field

3.1. Following Procedure and Checklists

Before going to the field, we dispatched letters to each facility by courier (with delivery confirmation and recipient contact information requested). Regular postal delivery was too slow and unreliable. Our letter briefly described the study, the data we were collecting, and that we requested an interview to complete a questionnaire. If a facility administrator was not available, we asked that a suitable substitute be identified. We sent e-mails with the same content prior to the letters, but there was little evidence the e-mails served to alert the facilities to our visit. While in the field, team members used mobile phones to call a day or two before arriving at facilities to confirm the availability of the facility administrator and HMIS data personnel.

In the field manual, we provided checklists for facility entry and exit, to ensure teams had everything they needed going in and that they left their sites with everything required for the study. We reprinted these single-page checklists for ready reference in the vehicles. We also provided a “night-before” homework checklist that included activities like charging batteries, replenishing the supply inventory, making arrangements with the driver, and communicating with headquarters for safety check-ins, progress reports, and schedule updates.

We prepared a script for teams to use as they entered facilities, and provided an official stamped and signed letter from the Ministry of Health authorizing access to HMIS forms. Data collectors made sure the questionnaires were completed only after obtaining signatures on institutionally approved, stamped, consent forms (with a copy to the facility).

At the conclusion of our visit, we paid a small fee in cash (Ugandan Shilling UGX 50,000, i.e., about USD 21) to compensate the organization for its staff time. We asked for a signed receipt, and indicated the funds should go to some sort of collective organizational effort

(e.g., a staff lunch). A fee of this sort is often expected in limited-resource settings. As the teams were preparing to leave, they reviewed their exit checklists and noted materials collected. This checklist also served to track follow-up actions, if any, needed to complete data collection. A second day visit to facilities was often required to complete data collection, and a visit to district offices was sometimes useful to obtain reports that were no longer available at the facility, but perhaps sent to the district office.

3.2. Managing Transportation, Accommodation, and Cash

One field team member was responsible to sign for cash advances, collect receipts, and settle up with the accounting staff upon return from the field. Teams were provided a modest lump-sum daily allowance for food and lodging (UGX 100,000), in addition to payment for their services (UGX 100,000). Generally, cash for both expenses and labor was advanced based on the number of days one was expected to be in the field; half the funds were provided in advance, and any remaining owed was paid upon return. Makerere established payment methods for fuel, and made plans in the event of repair or replacement of vehicle.

Makerere requested bids from several companies to provide drivers and vehicles for the six teams. We selected two companies, assigning three teams to each. Our team personnel were required to pay for petrol along the way, providing another accounting and cash flow challenge. In one region, it was necessary to hire security personnel from the Ugandan police force to accompany data collectors. Heavy rains made roads slippery, mud was deep, and some roads were washed out (see Figure 1).

3.3. Scanning Data Forms

We felt fortunate to be able to use paper form scanning technology in the field with considerable ease. Scanning source forms reduced field data capture time and provided high quality images that were easily retrievable for data entry and verification. We purchased seven Fujitsu ScanSnap S1500 scanning machines, thinking we might need a backup for one of our six teams, but we never did. These can scan source forms at 20 pages per minute, double sided, generating PDF files. Paper that was wet or mutilated could be enfolded in a clear plastic sleeve prior to scanning. The scanners provided good image results, even from faded or stained source documents.

In advance, we established electronic Dropbox folders on each computer by region, facility, and form type, so forms would be uploaded to uniformly-named files. The Dropbox company maintains a password-protected server copy of all files to serve as backup (although we also created DVD backups for each facility). Any invited individual can share access to the web-based folder. While small capacity Dropbox is free, we paid USD 10 monthly for each team and the data center to ensure access to 50 gigabytes of space on each account. Scanned files were stored to the laptop and automatically uploaded when internet service became available.

When electricity was not available at the facility, we requested use of any backup generator that was available, and provided an allowance to the facility for fuel costs. Alternatively, we set up a workspace near our vehicles (using our folding tables) and connected a power inverter to the car battery to operate our equipment (see Figure 2). We tried and rejected the

use of a separate deep-cycle 12 Volt battery. We purchased such batteries for each field team, but without out-of-car battery chargers, they went unused. Teams did not like handling the large heavy batteries, and were concerned about the battery acid spilling and damaging cars or clothing.

3.4. Maximizing Internet Service

Field teams used portable internet modems that plugged into the computers. Internet service was intermittent and slow for the high volumes of forms we were transmitting daily. We attempted to determine in advance whether one or another of the available service providers was more effective in some regions than others, but in the end had to provide two brands of modems to each team.

3.5. Software Selection and Enhancing Data Entry Accuracy

We provided an orientation to data entry for all personnel in the project, whether they were planning to be deployed to the field or intended to enter data in the headquarters office. Field personnel needed to understand the challenges faced by their team mates entering the data. We used the data entry software, *CSPro* (Census and Survey Processing System), which was freely available from the US Census Bureau. The attractive features included a strong *data dictionary* that allows range setting, skip and fill rules, and other conditional data checks triggered during data entry. Perhaps most important, *CSPro* has a data verification module that provided realtime verification by immediately notifying the data entry person when the second entry did not agree with the original. This eliminated the extra step of locating source document when the first and second entry are compared in the traditional method. We enhanced accuracy by designing data entry screens to resemble source documents.

During a first round of data collection from district offices, we conducted direct data entry from forms to the laptop computer in the field. The error rate from direct data entry in the field was unacceptably high, thus we switched to centralized data entry to provide better supervision and more efficient communication across data entry staff, all working in the same room.

3.6. Data Collected

We successfully collected, entered, cleaned and analyzed a large dataset of nationally representative health systems variables, collected directly from 315 health service facilities. We have 6 years of values for 670 variables from four primary data sources, for a total of 1.9 million data elements. We collected 87 per cent of the possible HMIS monthly values from facilities, and conducted interviews of administrators in 99 per cent of the facilities we visited, see Table 2.

4. Data Management Effort

We set up a data management center at Makerere University to handle incoming data, track completeness of requested forms, and enter data into electronic databases. A flipchart at the front of the room tracked progress. We provided six data entry staff (one for each field team) with laptop computers, with keyboard, mouse, and two screens—one for viewing the

scanned form images and the other for data capture screens. A supervisor and a data manager were also housed in the data management center, and an information technology specialist occasionally came by to troubleshoot internet, hardware, and software problems. Data entry staff downloaded information from their team's Dropbox folders, checked for quality (they completed daily data quality report forms), and compared expected versus received page counts for each facility. In retrospect, we should have deployed more data entry personnel from the start of the project, and provided them a larger office space. A larger data entry team could have provided faster data entry completion, better feedback to field teams concerning data quality, and a shorter lag between data collection and analysis.

Poor quality and missing data are often a significant problem in retrospective data collection exercises, especially in limited-resource settings; this project was no different. As we expected, there were more missing data points early in the observation period, 2005–2006, when the HMIS national system was newly implemented. A significant challenge in every facility we visited was the lack of proper storage for paper forms. Following completion of double-entry data verification, we selected a small random sample from each data set to compare directly with scanned source forms. From this analysis, we found fewer than 3 per cent of entries contained data entry errors.

Some facilities could not provide all the reports we sought, and some responses were missing from reports available. We collected catchment area population numbers from all but 13 per cent of facilities, and conducted additional research to complete the missing values. We wanted catchment area population data so we could conduct population-at-risk analysis. For example, if a facility monthly report did not contain maternal death data, the population for that facility-month would be excluded from analysis.

We took steps to ensure data quality and employed analytical methods to reduce bias. Even so, some variable values in the data set were clearly unreasonable. We tried several methods to clean outlier values, such as the median absolute deviation, identifying values that were two standard deviations from the mean, and others. Finally, we used graphical methods to detect outlier values. We graphed each variable for each facility over the 72 months of the observation period. Using this method, we spotted instances of unreasonable fluctuations in services from month-to-month. When values were missing or unreasonable, we compared our data with the value on the original scanned form. If the value on the source form was unreasonable, we treated that data item as missing in the analysis dataset. All this took time.

We underestimated the amount of time required for data entry and cleaning, which resulted in delays between collection and analysis. First and second data entry were completed by eight people working full time for 10 months. Cleaning of outliers and evaluating data entry accuracy required the full time work of one person for 3 months. We used *Stata* (Version 11) for most of the analysis and software was provided to partners as needed. Through an iterative process, in person, by e-mail, Skype, and phone, we conducted and refined analytical questions and details concerning which variables to use, potential confounders, and suitable statistical methods.

5. Producing and Publishing Manuscripts

To conceptualize manuscript topics, we conducted a series of meetings where partners discussed specific questions to be addressed, methods for analysis, and final interpretation of results. We found that submitting conference abstracts was a good way to focus attention and produce short-term products.

Each key team member assumed responsibility for at least one manuscript (we had a list of 12 topics) and was responsible to write the first draft. When all data had been collected and manuscripts drafted or at least outlined, key team members gathered for a week in Entebbe, Uganda to agree on interpretation of the findings. Some findings were other than expected, such as PEPFAR investments that were associated with small declines in outpatient care for young children, TB tests, and in-facility deliveries. This resulted in additional time evaluating alternative assumptions and revising manuscripts to address all concerns.

When the initial manuscript was drafted, it was submitted for clearance from the US Centers for Disease Control and Prevention (CDC), as required by CDC policy (Policy number CDC-GA-2005–06, *Clearance of Information Products Disseminated Outside CDC for Public Use*). The purpose of clearance is to ensure “all information products authored by CDC staff members or published by CDC and released for public use are of the highest quality and are scientifically sound, technically accurate, and useful to the intended audience.” The clearance process took approximately a year and involved more than a dozen CDC staff (Hagopian, Stover, & Barnhart, 2015).

6. Discussion

We observed several benefits of data collection with smartphones that have been previously reported, like the management of field teams through real-time data collection, and ensuring the authenticity of data (Tomlinson et al., 2009). Snapshots of field team members at each health facility’s road sign provided us with the geographical location, date, and time of arrival.

We believe it was more efficient to collect images of the original data forms at each facility, and then enter data from the conveyed images at a central location. The scanners, internet, and Dropbox provided the ability to upload images of forms remotely and convey them to the central data entry location at Makerere University where data collection could occur before teams returned from the field.

We developed electronic data capture screens for the administrator questionnaire, but chose paper forms instead. The time required to use electronic data capture has been reported to be shorter than for paper forms, however, data collectors felt less connection to the respondent during in-person interviews, in part because of diminished eye-to-eye contact, compared to using paper forms (King et al., 2013). Our teams found the text-heavy interview was quicker to complete on paper and we wanted to limit expensive field time.

7. Conclusion and Lessons Learned

This article describes the experience of conducting a public health program evaluation in Uganda. We offer our observations after visiting 315 health facilities, collecting 6 years of medical services data and completing structured interviews of health facility administrators.

Detailed and flexible organization was necessary to surmount the problems that arose while doing research in this limited-resource setting, especially in rural areas. A field manual was useful for maintaining consistency in data collection while managing the unexpected.

Lessons learned include the following:

- a. Throughout the project, our top priority was developing and maintaining collegial relationships. Navigating large organizations and government agencies is time intensive, and requires patience and humor. Individual relationships can rise above organizational annoyances (e.g., delay in funding, human research ethics requirements, and staff changes). We held a welcome meeting for data entry personnel to introduce people from all organizations and an end of project celebration and debriefing event. These served to build community and develop relationships for this project and for future opportunities.
- b. The formal data user agreement proved to be an important legal document. It signaled we took our Ministry of Health partner seriously as the owner of the data.
- c. There were many opportunities for building local research capacity while conducting this evaluation project. We employed students, provided training to junior faculty, and otherwise sought out opportunities to involve less experienced professionals.
- d. Identifying the titles and authors of manuscripts early in the project ensured each area of inquiry had a champion and minimized any possibility of future conflicts.
- e. Local institutions had very little working capital and little flexibility to absorb unexpected costs.
- f. Our investment in developing a thorough training manual aided the project in several ways. Developing the manual helped the principals understand what was expected of our teams and communicated the same information to everyone. The manual also aided data consistency by serving as a common source for data collection methods.
- g. Piloting our data collection instruments provided important feedback on how we approached health facility personnel, the format of our questions, and the availability of the data we wanted to collect.
- h. Hiring the right people and assigning them to appropriate roles was critical. Our data collectors were resourceful, talented, and innovative, willing to work very hard across extended hours and through rugged terrain to meet schedules. Mid-level management staff were equally effective.

- i. Fiscal management and cash controls were key to project accountability and managing a tight budget. The large number of expenses to keep six teams in the field (per diem, gas, lodging, internet access, phone service, and drivers) required a complex financial system. UW and Makerere people familiar with each institution's financial requirements created systems to track and monitor expenses and transfers.
- j. We found vehicle drivers were critical team members. They provided local knowledge of the road networks, available facilities for food, lodging, and vehicle maintenance, and provided security for the vehicle and the field team.

The importance of maintaining collegial relationships within and between partner organizations on this project cannot be overstated. Each of us worked for large organizations (each with its own accountability obligations), and we needed to ensure our own organizations did not get in the way of progress. From time to time, we also needed to help our partners maintain their own strong standing within their organizations by complying with their regulations. Three rounds of staff turnover in one of the organizations posed a significant challenge. Our collegiality was bolstered by friendship, humor, trustworthiness, and individual competence.

The US team members were grateful to learn the value of traditional welcome events as we entered health facilities, involving tea and preliminary conversations about the journey. These meetings provided an opportunity for teams to share food while becoming familiar with each other and obtaining the approval of facility leaders.

Large data collection efforts in limited-resource settings are not uncommon, but researchers have not published much on how they have managed these projects. We hope our detailed account will be useful for other researchers.

Acknowledgments

We would like to acknowledge the support for this work from several others including, Francis Runumi, Emily Bancroft, Jessica Crawford, and Nathaniel Lohman.

Funding was provided by the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) through the Centers for Disease Control and Prevention, under the terms of the public health evaluation project entitled "Assessment of the Impact of PEPFAR/Global Disease Initiatives on non-HIV Health Services and Systems in Uganda" (CE.08.0221).

The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

References

- Carletto G. Building panel survey systems in Sub-Saharan Africa: The LSMS experience [Audio file]. Speech delivered at the Symposium on Cohorts and Longitudinal Studies (Panel 6: The practicalities of cohort and longitudinal research, Moderator: Linda Adair) 2014 Retrieved from <https://www.unicef-irc.org/knowledge-pages/Symposium-on-Cohorts-and-Longitudinal-Studies-2014/1097/>
- Gladwin J, Dixon RA, Wilson TD. Implementing a new health management information system in Uganda. *Health Policy and Planning*. 2003; 18(2):214–224. Retrieved from . DOI: 10.1093/heapol/czg026 [PubMed: 12740326]

- Hagopian A, Stover B, Barnhart S. CDC clearance process constitutes an obstacle to progress in public health [Letter]. *American Journal of Public Health*. 2015; 105(6):e1. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4431081/>.
- Hagopian A, Flaxman A, Takaro TK, Al-Shatari SAE, Rajaratnam J, Becker S, Burnham G. Mortality in Iraq associated with the 2003–2011 war and occupation: Findings from a national cluster sample survey by the University Collaborative Iraq Mortality Study. *PLoS Medicine*. 2013; 10(10):e1001533. Retrieved from . doi: 10.1371/journal.pmed.1001533 [PubMed: 24143140]
- King JD, Buolamwini J, Cromwell EA, Panfel A, Teferi T, Zerihun M, Emerson PM. A novel electronic data collection system for large-scale surveys of neglected tropical diseases. *PloS ONE*. 2013; 8(9):e74570. Retrieved from . doi: 10.1371/journal.pone.0074570 [PubMed: 24066147]
- Lohman N, Hagopian A, Luboga SA, Stover B, Lim T, Makumbi F, Pfeiffer J. District health officer perceptions of PEPFAR's influence on the health system in Uganda, 2005–2011. *International Journal of Health Policy and Management*. 2017; 6(2):83–95. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5287933/>.
- Luboga SA, Stover B, Lim TW, Makumbi F, Kiwanuka N, Lubega F, Hagopian A. Did PEPFAR investments result in health system strengthening? A retrospective longitudinal study measuring non-HIV health service utilization at the district level. *Health Policy and Planning*. 2016; 31(7): 897–909. [PubMed: 27017824]
- Ministry of Health Uganda, ICF International, CDC Uganda, USAID Uganda, & WHO Uganda. Uganda AIDS indicator survey 2011/2012Aug. Retrieved from <https://dhsprogram.com/pubs/pdf/AIS10/AIS10.pdf>
- MittiR. Using Surveybe to improve the collection of panel data: Lessons from the Kagera Tanzania [Audio file]. Speech delivered at the Symposium on Cohorts and Longitudinal Studies (Panel 6: The practicalities of cohort and longitudinal research, Moderator: Linda Adair)2014Retrieved from <https://www.unicef-irc.org/knowledge-pages/Symposium-on-Cohorts-and-Longitudinal-Studies-2014/1097/>
- OpenDataKit. About(n.d.)Retrieved from <https://opendatakit.org/about>
- Tomlinson M, Solomon W, Singh Y, Doherty T, Chopra M, Ijumba P, Jackson D. The use of mobile phones as a data collection tool: A report from a household survey in South Africa. *BMC Medical Informatics and Decision Making*. 2009; 9 Article 51. Retrieved from . doi: 10.1186/1472-6947-9-51
- Uganda Bureau of Statistics & ICF International. Uganda demographic and health survey 2011/2012Aug. Retrieved from <http://www.ubos.org/onlinefiles/uploads/ubos/UDHS/UDHS2011.pdf>
- U.S. State DepartmentThe president's emergency plan for AIDS relief: Capacity building and strengthening framework (Version 2.0)Office of U.S Global AIDS Coordinator & Bureau of Public Affairs, U.S. State Department; 2012Retrieved from <https://www.pepfar.gov/documents/organization/197182.pdf>



Figure 1.
Field data collection teams faced transportation challenges. (Field teams were assisted by helpful local people in bad weather and bad road conditions.)

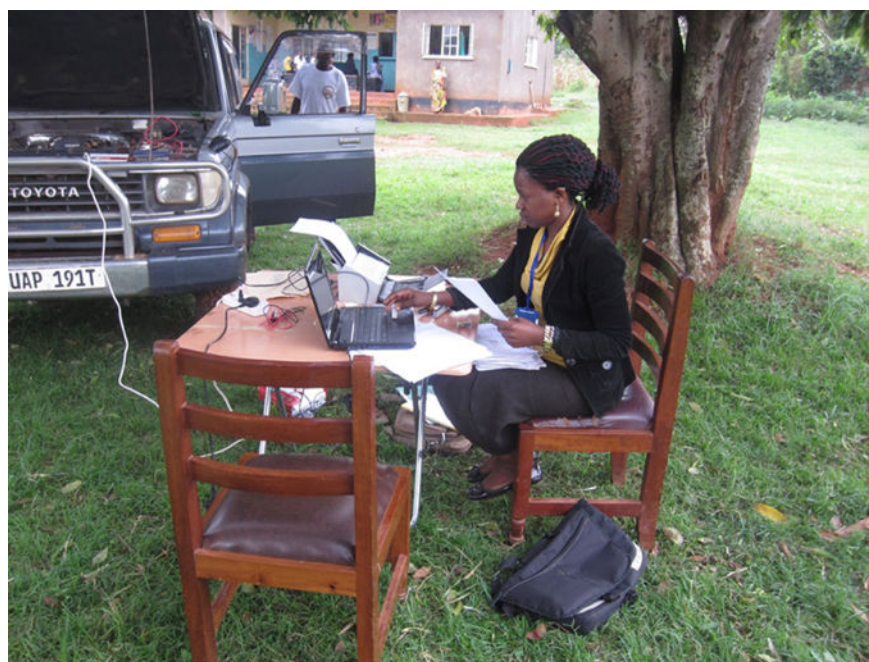


Figure 2.
Field data collection workspace powered by car battery. (Stellah Kayongo using folding table, laptop, and scanner in Bowa, Uganda, August 2012.)

Table 1

Uganda Public Health Evaluation Training Manual, 2012

Section	Contents	Appendices
Introduction	Who we are, who you are, why this study	
Study Description	Study background and goals	
Study Design	Healthcare facility selection, data collection forms, identification of forms to scan, identification of variables for data entry, other data sources used for this study Description, example screenshots, and introduction to <i>CSPro</i> (Census and Survey Processing System) data entry software	<ul style="list-style-type: none"> List of targeted facilities Facility questionnaire Pharmacy log HMIS forms Document scanner instructions CSPro data entry instructions
Notifying Facilities	Advance letter sent by courier Instructions for calling facility in advance, scheduling an appointment, introducing the team	<ul style="list-style-type: none"> Introduction letter including legal authority
Rapport and Attitude	Building a relationship, participation is voluntary, consent for participation, privacy, efficient interviewing, housekeeping	<ul style="list-style-type: none"> Introductory script
Scenario at the Facility	Expected order of events, team photograph, GPS reading, ODK smartphone data completion, scheduling with facility administrator, obtain assistance from facility data manager Instructions for completing the interview, use pre-named folders and files, instructions for uploading data from laptops, backup laptop data to DVD	<ul style="list-style-type: none"> ODK facility arrival form ODK facility exit form Instructions for using ODK Receipt for payment to facility Entry and exit checklist Data quality report Photography consent form
Consent Process	Who is consenting, consent process script, documentation, confidentiality of data	<ul style="list-style-type: none"> Consent form
Scheduling	Dates and times of training and data collection	
Roles and Responsibilities	Team structure and member roles, monitor supplies, communication with data center and administrator, data entry progress report, data backup procedures	<ul style="list-style-type: none"> Daily homework checklist Data quality and integrity contract
Training Review	Expectations for training completion, demonstration of skills, logistical organization and compensation	<ul style="list-style-type: none"> Training evaluation form
Contact Information	Study personnel, field and data team members, individual photographs	<ul style="list-style-type: none"> Study contact information and emergency contacts
Supplies and Equipment	Supply inventory, field manual, extra printed forms, financial accounting materials, laptop computer, scanner, smartphone, modem, 12 Volt power inverter, table and chairs	<ul style="list-style-type: none"> Assigned facilities list with contact information and map
Problem Resolution	Computer corruption prevention, problems with electronics, car battery as power source, lost equipment, car breakdown, emergencies	
Travel and Logistics	Team deployment and scheduling visits to each facility	<ul style="list-style-type: none"> Route and schedule worksheet

Table 2

Data Collection Results

Data Source	Number of Forms Possible	Number of Forms Obtained	Number of Variables	Data Capture Method
Facility HMIS monthly reports	22,680	19,725	84	Scan of paper forms
Facility HMIS annual reports	1,890	783	183	Scan of paper forms
Health Facility Survey	315	312	381	Interviews completed on paper
Pharmacy Log	945	877	22	Review of pharmacy records